Sci. Rep. Kanazawa Univ.
Vol. 36, No. 1 pp. 21-29
July 1991

Evidence for Archaetypal Code ?

Keishiro WADA

Department of Biology, Faculty of Science, Kanazawa University

Abstract : Bacterial ferredoxins, especially ferredoxins from *Clostridia,* fermentative bacteria, are very primitive proteins in which the characteristics of the organisms in early evolutionary time are retained. The very limited existence of amino acids specified by the third base of codons in the present genetic code is found in clostridial ferredoxins. This may be evidence for the archaetypal genetic code. Clostridial ferredoxins are constituted from simple amino acids which are specified by the first two bases of codons, and either one of the amino acid pairs in weak codon family (LARGERKVIST, 1978), except for GAN-codon family for glutamic acid and aspartic acid.

It is proposed on the primitive coding system of early evolutionary time that the species of amino acids used have been less than sixteen and that the first two bases of triplet codons have specified amino acids and the third base has functioned only in frame-deciding.

1. Bacterial Ferredoxins

Bacterial ferredoxins are the small redox proteins with one or two [4Fe-4S] cluster(s) as an active center. Thus far, ferredoxins from over 20 different bacteria were sequenced (MATSUBARA *et al.*, 1980; SATO *et al.*, 1981). ADMAN *et al.* (1973) established the tertiary structure of *Peptococcus aerogenes* ferredoxin by x-ray crystallography. From comparison of bacterial ferredoxin sequence available, clostridial ferredoxins seem to be a fundamental form in terms of the distribution of eight cysteine residues which are important in holding iron-sulfur clusters (MATSUBARA *et al.*, 1980). It is remarkable that ferredoxin from *Clostridium butyricum* has no complicated amino acid residues such as arginine, lysine, histidine, tyrosine, tryptophan, methionine and leucine (Table 1), and that it consists of simple amino acids which have been found to occur in the product obtained by heating mixture of simple gases such as methane, ammonia and cyanide. On the other hand, plant-type (spinach) ferredoxin consists of all protein-constituting amino acids except for methionine.

Table 1 Amino acid compositions of ferredoxins

	1	2	3	4
Asp	5	5	2	11
Asn	4	3	1	2
Thr	3	1	4	8
Ser	3	5	2	7
Glu	2	2	7	9
Gln	3	2	1	4
Pro	3	3	4	4
Gly	6	5	5	6
Ala	6	7	12	9
Cys	8	8	9	5
Val	6	6	4	7
Met	0	0	0	0
Ile	4	5	5	4
Leu	0	0	1	8
Tyr	0	0	3	4
Phe	2	2	1	2
Lys	0	1	0	4
His	0	0	0	1
Arg	0	0	0	1
Trp	0	0	0	1
Total	55	55	61	97

1: Clostridium butyricum, 2: C. pasteurianum, 3: C. thiosulfatophilum, 4: spinach

TANAKA *et al.* (1964) reported that there was an internal repetition in the sequence of clostiridial ferredoxins. MATSUBARA *et al.* (1980) described that a proto-type of ferredoxins with iron-sulfur cluster(s) (MATSUBARA *et al.*, 1968) might be synthesized in early evolutionary time by using iron-and sulfur-compounds present probably in abundance on the primordial earth. Therefore, it is plausible that bacterial ferredoxins must be one of the proteins preserving the characteristics of ancient organisms which appeared on the primordial earth. They may conceal many of old interesting information in themselves on the evolutionary point of view.

2. Archaetypal Genetic Code

It was around 1966 that the genetic code was deciphered experimentally, mainly using systems from *Escherichia coli*. Since, the genetic code has been considered to be common to all organisms. Such universality of the genetic code was broken by finding that vartebrate mitochondria have minor changes in their code table. Evolution of the genetic code has been discussed by many investigators (CRICK, 1968; ORGEL, 1968; JUKES, 1973; HARTMEN, 1975; WONG, 1975; LARGERKVIST, 1978; HASEGAWA and MIYATA, 1980). Early organisms could not use the complicated amino acids, before these amino acids became available or their synthetic pathways were acquired (JUKES, 1968). Each of simple amino acids used by primitive organisms have

Table 2 Proposed Archaetypal Code and Subsequent Evolution

Codon Family	Early Genetic Code (in this paper)	Present Genetic Code	Archaetypal Assignments (from T. Jukes, 1968)	Present Assignments
UUN			Phe	UUY-Phe UUR-Leu
CUN			Leu	No change
AUN			Ile	AUY-Ile AUA-Ile AUG-Met
GUN			Val	No change
UCN	a)		Ser	No change
CCN			Pro	No change
ACN			Thr	No change
GCN			Ala	No change
UAN			Tyr	UAY-Tyr UAR-C. T.
CAN	Gln[b]	CAY-His	His	CAY-His
		CAR-Gln		CAR-Gln
AAN			Asn	AAY-Asn AAR-Lys
GAN	Asp/Glu[c]	GAY-Asp	Asp	GAY-Asp
		GAR-Glu	Glu	GAR-Glu
UGN			Cys	UGY-Cys UGA-C. T. UGG-Trp
CGN			Arg	No change
AGN			Ser	AGY-Ser AGR-Arg
GGN			Gly	No change

N= A,G,C,or U; Y= U or C; R= A or G; C. T. = chain termination. a) The not-described are the same as the proposal by Jukes (1968). b) CAN codeis proposed to be correspond to Gln in the early genetic code. c) GAN in the early genetic code was read without discrimination between Asp and Glu.

four codons arranged in a codon family (or a quartet by JUKES, 1968) such as GGU, GGC, GGA and GGG for glycine and GCU, GCC, GCA and GCG for alanine. The wobbling pairs between the first base of anticodons and the thrid base of codons such as U-A, U-G, G-U, G-C, I-U, I-C and I-A can occur (CRICK, 1966) and many modified bases in the first base of anticodons are found at the present. These facts lead us to some interesting speculations on evolution of the genetic code.

One of them is that the early organisms have adopted 'doublet code' rather than triplet code. The doublet code may not imply that DNA consists of the consecutive doublet but that the first two bases of codons are important in specifying amino acids and the third base is used as a sort of spacer to decide the codon frame. This speculation can be supported by that mammalian mitochondrial code differs in some respects from the universal genetic code (BARREL *et al.*, 1980; JUKES, 1981).

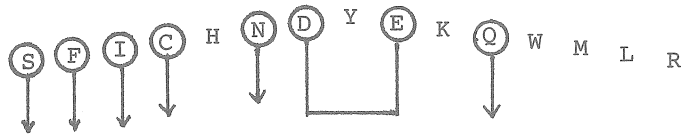
JUKES (1968) proposed archaetypal genetic code which composed of not more than 16 families (or quartets) as shown in Table 2. Eight families (UCN, CUN, CCN, CGN, ACN, GUN, GCN, and GGN) of them were retained in the present genetic code. These are codon families with strong G-C type codon-anticodon interaction in the first two positions of codons (LARGERKVIST, 1978). On the other hand, the remaining codon families with weak A-U type interaction were, during evolution, devided into purine-terminated and pyrimidine-terminated codons, the respectives of which were applied toward two different amino acids available newly. However, to obtain the experimental evidence for the archaetypal code is nearly impossible.

3. Graphical Comparisons of Amino Acids Coded by Weak Codons in Bacterial Ferredoxins

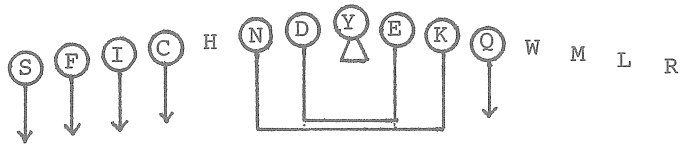
As mentioned above, amino acid compositions of clostridial ferredoxins are very characteristic. In order to stress this, especially focusing amino acids coded by weak codons, graphical comparisons were carried out as shown in Fig. 1. Amino acid pairs in the weak codon families were located one by one in both sides of tyrosine. Amino acids coded by pyrimidine-terminated codons were arranged in the left side of tyrosine (Y) and ones by purine-terminated codons, in the right side. If either one of amino acid pairs in the weak codon family is contained in a certain protein, the sign of amino acid is circled and marked by arrow. On the other hand, if both in the weak codon family are present in it, the signs of amino acids are also circled and two are connected by line. When tyrosine is contained in the protein, it is circled and marked with △ as a case that both of amino acid pairs are present, as tyrosine and termination codons share UAN-codon family.

As shown in Table 1, seven species among amino acids naturally found in protein at the present day were lacked in the case of *C. butyricum* ferredoxin. The connected with a line was only a pair of glutamic acid and aspartic acid. The other amino acids such as asparagine, glutamine, cysteine, isoleucine, phenylalanine and serine (When considered as

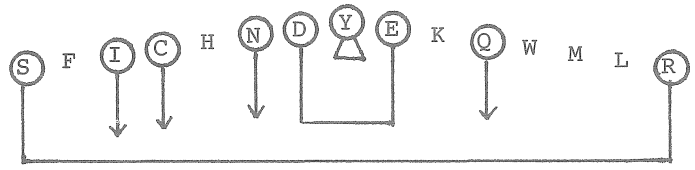
Clostridium butyricum Fd (55)



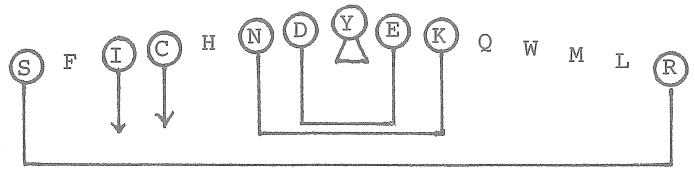
C. pasteurianum Fd (55)



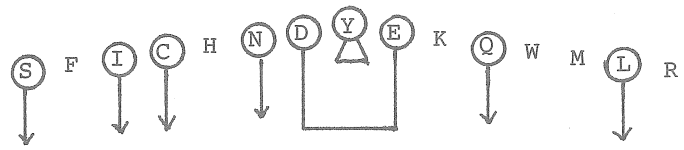
C. acidi-urici Fd (55)



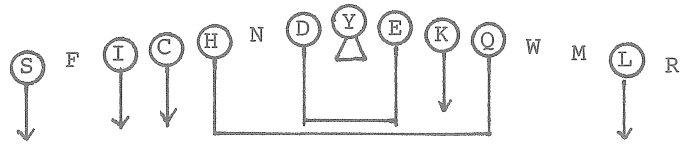
C. M-E Fd (55)



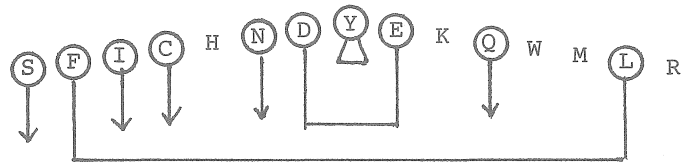
Chlorobium limicola Fd (60)



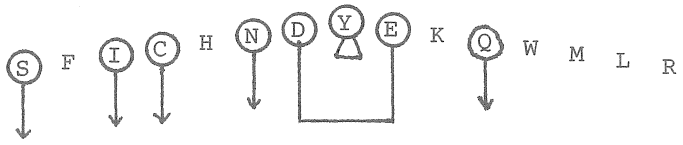
Clostridium tartarivorum Fd (55)



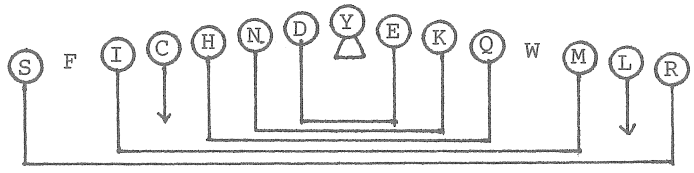
C. thiosulfatophilum Fd (61)



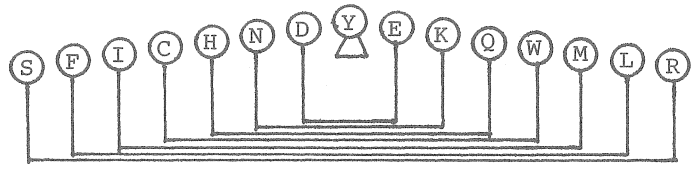
Ancestor Fd (29)



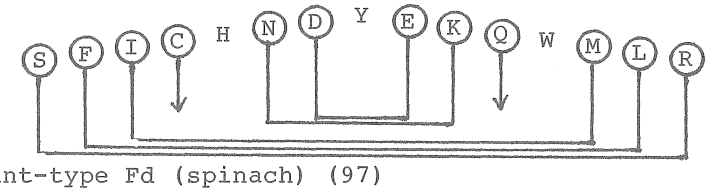
Chromatium vinosom Fd (82)



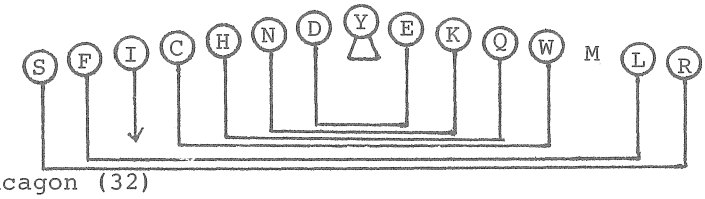
Mycobacterium smegmatis Fd (106)



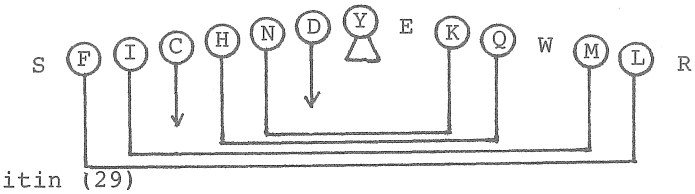
Desulfovibrio gigas Fd (56)



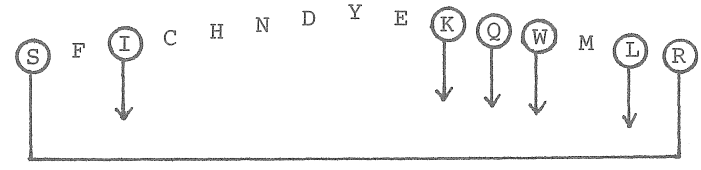
Plant-type Fd (spinach) (97)



glucagon (32)



melitin (29)



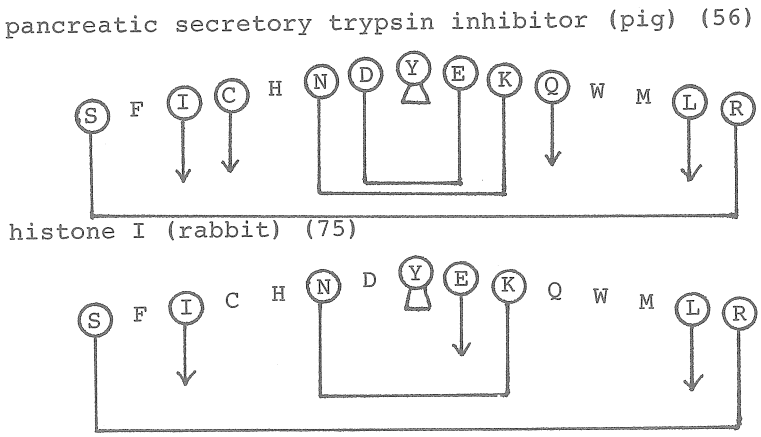


Figure 1

Graphic comparisons of amino acid pairs in the weak codon families.
Amino acid pairs in the weak codon families were located one by one in both sides of tyrosine (Y). Amino acids coded by pyrimidine-terminated codons were arranged in the left side of tyrosine and ones by purine-terminated codons, in the right side. If either one of amino acid pairs in the weak codon family is contained in a certain protein, the sign of amino acid is circled and marked by arrow. If both in the weak codon family are present in it, the signs of amino acids are also circled and connected by line. When the protein contains tyrosine, the sign of tyrosine is circled and marked with △ as a case that both of amino acid pairs are present. An indication of amino acids was by one letter notation.

the weak codon family) were marked with arrows. These facts mean that this ferredoxin has either one of amino acids belonging to the weak codon families. The same phenomena are found in the other clostridial and chlorobial ferredoxins and also in the ancestor ferredoxin discussed as benchmark in the early evolution of ferredoxin molecule (MATSUBARA *et al.*, 1968). *Mycobacterium smegmatis* and plant-type (spinach) ferredoxins contain almost all of protein-constituting amino acids and show the symmetrical graphical patterns. The further interesting is that most of amino acids present in ferredoxins from *Clostridia* and *Chlorobia* are located in the left side of tyrosine in the graphical comparisons in Fig. 1. The fact shows that they consist of the amino acids coded by pyrimidine-terminated codons with the exceptional case of glutamine and histidine pair as discussed later. Some of small other proteins were also compared in this way. As shown in Fig. 1, glucagon, melitin, pancreatic secretory trypsin inhibitor (pig) and histone I (rabbit) (DAYHOFF, 1972) did not show any regularities in their graphical patterns.

4. Discussion

Ferredoxins, including the other iron-sulfur proteins, seem to be of the most fundamental proteins, which first acquired a redox function in the early organisms. This is readily

evidenced from that the iron-sulfur proteins have a quite ubiquitous distribution ranging from primitive prokaryotes to all living cells of eukaryotes on the present earth. The early genetic code considered from the amino acid usage in bacterial ferredoxins is in good agreement with the archaetypal code proposed by JUKES (1968). Amino acids by the pyrimidine-terminated codons among the weak codon families in the present genetic code seem to be more primitive than those by the purine-terminated codons. With increase of the number of amino acid species used as protein constituents during evolution, the purine-terminated codons became to be devoted to the 'new' amino acids and termination codons in some cases. However, there are remained many ambiguities and not-answered questions in those results mentioned above. As all of clostridial proteins do not necessarily show the uneven usage of amino acids, these may be events governing by contingency.

There are many cases that both of glutamic acid and aspartic acid are containing in the 'primitive' ferredoxins. At the present glutamic acid is coded by the purine-terminated codons (GAA and GAG). According to JUKES (1968), codons for these two amino acids owing to their similarities may have not been distinguished in the primitive proteins but used interchangeably.

Another question is the histidine-glutamine pair in CAN-codon family. It is not sure if histidine has been contained in the crude tarry mixtures formed by heating and by electric discharge in the mixture of simple gases. While, glutamine is synthesized from glutamic acid and ammonia by the ATP dependent enzyme reaction. Thereby, Jukes let CAN-codon family correspond to histidine in the archaetypal code. Is it not plausible that CAN-codon family has coded glutamine in the early genetic code and that the pyrimidine-terminated codons have been applied toward the 'new' amino acid, histidine ? Because asparagine is assigned in the archaetypal code, AAN-codon family. It is also a question why the purine-terminated codons of CAN-codon family has not been converted to the codons of a 'new' amino acid, histidine, like in almost all of the weak codon family, when histidine was lately added to amino acids as protein constituent. Asparagine is also synthesized by the same manner as glutamine. Glutamine is essential as a nitrogen source to tryptophan, histidine and the other important building blocks in the present metabolic pathway. Histidine residue is found less often in bacterial ferredoxin molecules. It is the case in the most of proteins. Furthermore, there is no report, even though histidine residue(s) exist in plant-type ferredoxins, that the role of important biological function has been ascribed to histidine residue(s). In the primordial earth, there may have been the way to synthesize glutamine or asparagine abiotically. Therefore, I would like to assign glutamine, instead of histidine, to CAN-codon family in the early genetic code.

The reasons are not clear why serine, arginine and leucine have six each codons in the present genetic code. CGN-codon family for arginine is proposed to have been read as ornithine in the early evolutionary time (JUKES, 1974). This proposal is very interesting when we consider a hypothesis that the strong codon families are assigned to simple amino acid with the low molecular weight (HASEGAWA and MIYATA, 1980) as ornithine has

rather small molecular weight. We have no evidence for this at all. As shown in Fig. 1, *C. acidi-urici* and *C. M-E* ferredoxins have both serine and arginine. However, these amino acids may not share AGN-codon family but may be coded by the separate strong codon families, UCN-for serine and CGN for arginine. In the same way, leucine in *C. tartarivorum* and chlorobial ferredoxins may not be coded by the purine-terminated codons, UUR (UUA and UUG), but by the strong codon family, CUN.

At present, we have not-enough data of ferredoxin gene sequences available to be compared. Evidence for the archaetypal code mentioned above may become more clear when we have gene sequences of clostridial and chlorobial ferredoxins in the near future.

References

ADMAN, E. T., L. C. SIEKER and L. H. JENSEN (1973) The structure of a bacterial ferredoxin. *J. Biol. Chem.*, **248**, 3987-3996.

BARREL, B. G., S. ANDERSON, A. T. BANKIER, M. H. L. de BRUIJN, E. CHEN, A. R. COULSON, J. DROUIN, I. C. EPERON, D. P. NIERLICH, B. A. ROE, R. STADEN and I. G. YOUNG (1980) Different pattern of codon recognition by mammalian mitochondrial tRNA. *Proc. Natl. Acad. Sci. USA*, **77**, 3164- 3166.

CRICK, F. H. C. (1966) Codon-anticodon pairing: the wobble hypothesis. *J. Biol. chem.*, **19**, 548-555.

----- (1968) The origin of the genetic code. *J. Mol. Biol.*, **38**, 367-379.

DAYHOFF, M. O. (1972) Atlas of Protein Sequence and Structure, Vol. 5, National Biomedical Research Foundation, Washington D.C.

HARTMEN, H. (1975) Speculation on the evolution of the genetic code. *Origin of Life*, **6**, 423-427.

HASEGAWA, M. and MIYATA, T. (1980) On the antisymmetry of the amino acid code table. *Origin of Life*, **10**, 265-270.

JUKES, T. H. (1973) Possibilities for the evolution of the genetic code from a preceding form. *Nature*, **246**, 22-26.

-----, (1974) On the possible origin and evolution of the genetic code. *Origin of Life*, **5**, 331-350.

-----, (1981) Amino acid codes in mitochondria as possible clues to primitive codes. *J. Mol. Evol.*, **18**, 15-17.

LARGERKVIST, U. (1978) "Two out of Three": an alternative method for codon reading. *Proc. Natl. Acad. sci. USA*, **75**, 1759-1762.

MATSUBARA, H., T. H. JUKES and C. R. CANTOR (1968) Structural and evolutionary relationship of ferredoxins. *Brookhaven Symp. Biol.*, **21**, 201-216.

-----, T. HASE, S. WAKABAYASHI and K. WADA (1980) Structure and evolution of chloroplast- and bacterial-type ferredoxins. *In*: *Evolution of protein structure and*

function, ed. D. S. SIGMAN and M. A. B. BRAZIER, Acad. Press, N. Y. pp. 245-266.

ORGEL, L. E. (1968) Evolution of the genetic apparatus. *J. Mol. Biol.,* **38,** 381-393.

SATO, S., K. NAKAZAWA, K. HON-NAMI and T. OSHIMA (1981) Purification, some properties and amino acid sequence of *Thermus thermophilus* HB8 ferredoxin. *Biochim. Biophys. Acta,* **668,** 277-289.

TANAKA, M., R. NAKANISHI, A. BENSON, H. F. MOWER and K. T. YASUNOBU (1964) The amino acid sequence of *Clostridium pasteurianum* ferredoxin. *Biochem. Biophys. Res. Commun.,* **16,** 422-427.

WONG, J. T. (1975) A co-evolution theory of the genetic code. *Proc. Natl. Acad. Sci. USA,* **72,** 1909-1912.